



SPOŁECZNO-EKONOMICZNE UWARUNKOWANIA ZACHOWAŃ WYBORCZYCH W WYBORACH DO SEJMU 2007 I 2011. BADANIE OPARTE NA DANYCH ZAGREGOWANYCH

*Michał Pierzgałski**

SOCIOECONOMIC DETERMINANTS OF VOTING BEHAVIOR
IN 2007 AND 2011 SEJM ELECTIONS.
ANALYSIS OF AGGREGATED DATA

— ABSTRACT —

In the paper I show the results of my research on voting behavior in Polish parliamentary elections. My studies focus on the elections to the Sejm in 2007 and 2011. Employing regression models I try to verify the following hypotheses: H1: An increase in voting turnout, that is an important indicator of positive social capital, is positively correlated with the Civic Platform election result, while increasing the negative social capital determinants results in the decline in the Civic Platform support. H2: An increase in voting turnout is beneficial for the Civic Platform, while having negative impact on the political support given to the Law and Justice and the Polish People's Party. H3: The higher the percentage of women in a county (powiat), the greater the average support for the Civic Platform. H4: When it comes to the determinants of political support for the selected parties, it turns out that, economic variables are statistically insignificant.

To carry out my studies I use statistical data available from Central Statistical Office of Poland and from National Electoral Commission.

— KEYWORDS —

voting behavior in Poland, regression models, socio-economic determinants of voting behavior

* Uniwersytet Łódzki, Instytut Studiów Politologicznych.

WPROWADZENIE

Celem badania, którego wyniki przedstawiono w artykule, była analiza wpływu wybranych zmiennych społeczno-ekonomicznych na zachowania wyborcze w Polsce na przykładzie wyborów do Sejmu RP w 2011 i 2007 roku. Wśród zmiennych, które zostały wzięte pod uwagę, znajdują się zmienne związane ze stanem gospodarki, jak również zmienne społeczno-demograficzne, a w tym determinanty tzw. kapitału społecznego¹.

Badając zachowania wyborcze, można wykorzystać zarówno dane indywidualne, które można pozyskać, przeprowadzając badania sondażowe, jak również dane na określonym poziomie agregacji (np. na poziomie gminy, powiatu). W artykule przedstawiono wyniki badania, które uzyskano w wyniku analizy danych zagregowanych na poziomie polskich powiatów (NUTS4) w latach 2007 i 2011.

Podstawowy zbiór wykorzystanych zmiennych został rozszerzony o dwie pomocnicze zmienne binarne oraz tzw. wskaźnik feminizacji: zmienna REGION – 1 dla powiatów wschodnich, 0 dla powiatów zachodnich; zmienna _It_2011–1 dla roku 2011, 0 dla roku 2007; oraz zmienna FEM – wskaźnik feminizacji (liczba kobiet w relacji do ogólnej liczby mieszkańców powiatu).

Na potrzeby badania przyjęto kilka hipotez, które poddano weryfikacji empirycznej: H1: Wzrost frekwencji wyborczej, która stanowi istotny wskaźnik pozytywnego kapitału społecznego, jest dodatnio skorelowany z wynikiem wyborczym Platformy Obywatelskiej (PO), podczas gdy zwiększenie wartości wyznaczników tzw. negatywnego kapitału społecznego (np. rosnący poziom przestępczości, bezrobocia) nie sprzyja PO. H2: Wyższa frekwencja wyborcza sprzyja PO i jest niekorzystna dla pozostałych partii politycznych objętych badaniem, a mianowicie dla Prawa i Sprawiedliwości (PIS) oraz Polskiego Stronnictwa Ludowego (PSL). H3: Im wyższy poziom feminizacji powiatu (im

¹ Kapitał społeczny można zdefiniować jako potencjał wynikający ze spistości relacji międzyludzkich. Kapitał społeczny jest wyrazem siły więzi społecznych, które służą interesom społecznym, a nie wyłącznie interesom grupowym lub indywidualnym. Zob. J.J. Sztaudynger, *Rodzinny kapitał społeczny a wzrost gospodarczy w Polsce*, „*Ekonomista*” 2009, nr 2, s. 192–193. Należy podkreślić, że kapitał społeczny nie jest zmienną, którą łatwo można zmierzyć. Niemniej można wskazać wiele czynników, które z różnym natężeniem wpływają na zasoby kapitału społecznego. O kapitale społecznym zob. R. Putnam, *Społeczny kapitał a sukces instytucji*, [w:] *Socjologia. Lektury*, red. P. Sztompka, M. Kucia, Kraków 2009, s. 388–396; P. Sztompka, *Zaufanie, nieufność i dwa paradoksy demokracji*, [w:] *Socjologia. Lektury*, op.cit., s. 397–408; R. Putnam, *Bowling Alone: The Collapse and Revival of American Community*, Nowy Jork 2000, ss. 541.

więcej kobiet w powiecie w stosunku do liczby mężczyzn), tym większe poparcie wyborcze udzielane PO. H4: Czynniki ekonomiczne, takie jak stopa bezrobocia lub średni dochód, nie są istotnie skorelowane z poziomem poparcia wyborczego udzielanego badanym partiom politycznym.

Wyniki badań zaprezentowano w postaci modeli regresyjnych dla danych panelowych, których parametry oszacowano z zastosowaniem tzw. zero-jedynkowej metody najmniejszych kwadratów (*Least Squares Dummy Variables*, LSDV).

Jak wspomniano, w artykule zostaną przedstawione modele ekonometryczne, który opisują charakter zależności między wybranymi zmiennymi objaśniającymi i poparciem wyborczym dla trzech partii politycznych w Polsce, a mianowicie dla PIS, PO i PSL. W artykule opisano również model regresyjny objaśniający wpływ wybranych zmiennych na poziom frekwencji wyborczej, która stanowi ważny wyznacznik kapitału społecznego².

Podstawową jednostką badawczą jest obszar powiatu (NUTS4). Liczba powiatów w Polsce w badanym okresie jest równa 379, a więc liczba obserwacji na wybranych zmiennych objaśnianych i objaśniających wynosi 758. Ze względu na stosunkowo mały rozmiar badanej populacji generalnej, jak również możliwość analizy każdego spośród jej elementów wyniki badań przeprowadzonych przez autora zostały osiągnięte wskutek analizy wszystkich jednostek populacji.

BADANIE ZACHOWAŃ WYBORCZYCH W POLSCE W WYBORACH DO SEJMU 2007 I 2011

W celu przedstawienia struktury zachowań wyborczych w Polsce należy mieć na uwadze trzy podstawowe problemy³.

Po pierwsze, zachowania wyborcze w Polsce cechuje przeciętnie stosunkowo znaczny poziom absencji wyborczej.

Po drugie, badania nad zachowaniami wyborczymi w państwach pokomunistycznych, w tym w Polsce, charakteryzują się dużym znaczeniem podziałów historycznych. Przede wszystkim decyzje podejmowane przez wyborców są m.in.

² Wszystkie modele regresyjne i diagramy przedstawione w tekście zostały wykonane przez autora w programie do obliczeń statystycznych STATA, na podstawie danych empirycznych dostępnych na stronach www Państwowej Komisji Wyborczej oraz w Banku Danych Lokalnych GUS (http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks).

³ J.J. Wiatr, *Europa pokomunistyczna*, Warszawa 2006, s. 228–242.

uwarunkowane występowaniem podziału socjopolitycznego, który określa się mianem podziału postkomunistycznego⁴.

Po trzecie, cechą charakterystyczną zachowań wyborczych w Polsce jest duża zmienność preferencji wyborczych. Poziom chwiejności wyborczej (*volatility*) w Polsce jest większy w porównaniu z państwami demokratycznymi Europy Zachodniej⁵.

Na wstępie rozważań na temat społeczno-ekonomicznych uwarunkowań zachowań wyborczych w Polsce warto krótko przedstawić charakterystykę polskiego systemu partyjnego w okresie 2007–2011.

Wyniki ostatnich wyborów parlamentarnych, m.in. elekcji w 2007 i 2011 roku, dowodzą, że możemy zaobserwować grupę partii politycznych, które utrwały swoją obecność na polskiej scenie politycznej. W tej grupie znajdują się: PO, PIS, PSL i SLD. Mechanizm działania polskiego systemu partyjnego wskazuje, że w przypadku Sejmu mamy do czynienia z systemem wielopartyjnym o umiarkowanej polaryzacji⁶.

Poparcie wyborcze dla partii, które przekroczyły 5-proc. próg wyborczy w wyborach do Sejmu w latach 2007 i 2011, przedstawia tabela 1.

Tabela 1. Poparcie wyborcze dla największych partii politycznych w wyborach do Sejmu 2007 i 2011

Lp.	Nazwa uczestnika wyborów	2007	2011
1.	Prawo i Sprawiedliwość (PIS)	32,1	29,9
2.	Platforma Obywatelska (PO)	41,5	39,2
3.	Polskie Stronnictwo Ludowe (PSL)	8,9	8,4
4.	Ruch Palikota	-	10,0
5.	Sojusz Lewicy Demokratycznej* (SLD)	13,2	8,2

* W wyborach w 2007 roku partie określane jako lewicowe, w tym SLD, utworzyły wspólny komitet wyborczy pod nazwą Lewica i Demokraci (LID).

Źródło: opracowanie własne na podstawie: PKW, <http://pkw.gov.pl/>, odczyt z dn. 20.06.2013.

⁴ M. Grabowska, *Podział postkomunistyczny*, Warszawa 2004, s. 207–278.

⁵ *Niestabilność wyborcza w Polsce*, red. M. Cześnik, Warszawa 2010, s. 9.

⁶ Wprowadzenie systemu większości względnej i jednomandatowych okręgów wyborczych w wyborach do Senatu RP przed wyborami w 2011 roku najprawdopodobniej doprowadzi do utrwalenia w przypadku izby wyższej systemu rywalizacji dwupartyjnej (PO i PiS).

Przedstawiona w artykule analiza zachowań wyborczych w latach 2007 i 2011 dotyczy trzech partii politycznych: PSL, PO i PIS, czyli partii, które stały się trwałym elementem polskiego systemu partyjnego i uczestniczyły ponadto w badanym okresie w koalicjach rządowych.

SPOŁECZNO-EKONOMICZNE UWARUNKOWANIA ZACHOWAŃ WYBORCZYCH W POLSCE

Jak wspomniano na wstępie, zbiór badanych zmiennych uwzględnia przede wszystkim determinanty stanu gospodarki oraz czynniki społeczno-demograficzne, które oddziałują na tzw. kapitał społeczny⁷.

Kapitał społeczny nie jest zmienną w pełni mierzalną. Jednak poziom kapitału społecznego możemy oszacować, analizując zmienne, które pośrednio lub bezpośrednio wpływają na jego wielkość.

Katalog wyznaczników kapitału społecznego nie jest zamknięty, dlatego w artykule sprawdzono oddziaływanie na poziom poparcia wyborczego tych wyznaczników, których wartości były łatwo dostępne na żądanym poziomie agregacji. I tak, wśród wskaźników bezpośrednio lub pośrednio oddziałujących na zasoby kapitału społecznego (negatywnego lub pozytywnego) można wymienić (kolejność przypadkowa) np.:

- 1) frekwencję wyborczą,
- 2) poziom przestępczości,
- 3) wskaźnik urodzeń,
- 4) stosunek liczby rozwodów do łącznej liczby rozwodów i małżeństw.

Wybór zmiennych, które będą przedmiotem analizy, uwarunkowany jest m.in. łatwością dostępu do wiarygodnych danych na odpowiednim poziomie agregacji: krajowy (NUTS 0), regionalny (NUTS 1), wojewódzki (NUTS 2), podregionalny (NUTS 3), powiatowy (NUTS 4), gminny (NUTS 5). Należy

⁷ R.D. Putnam wyróżnił kilkanaście mierników kapitału społecznego (Zob. np. J.J. Sztudynger, *Społeczne problemy wzrostu gospodarczego – analiza ekonometryczna*, [w:] *Etyka i ekonomia*, red. B. Klimczak, A. Lewicka-Strzałecka, Warszawa 2007, s. 132–133), a wśród nich: mierniki społecznej działalności w klubach i innych organizacjach, jak np.: liczba tych organizacji na 1000 mieszkańców, liczba spotkań klubowych; miary zaangażowania w sprawy publiczne: uczestnictwo w wyborach lub w zebraniach dotyczących spraw miejskich albo szkolnych; miary bezinteresownego zaangażowania społeczności: liczba organizacji *non-profit* na 1000 mieszkańców itp.; miary nieformalnego uspołecznienia (np. częstotliwość przyjmowania gości w domu); miary społecznego zaufania. Na podstawie tych mierników R.D. Putnam zbudował syntetyczny indeks kapitału społecznego.

podkreślić, że w artykule pominięto, z powodu trudności związanych z dostępem do danych na odpowiednim poziomie agregacji, kilka zmiennych, które mogłyby wzbogacić wyniki badania. Wśród tych zmiennych znajdują się m.in. poziom spożycia alkoholu czy stosunek aktywnych organizacji *non-profit* do liczby mieszkańców.

OGÓLNA CHARAKTERYSTYKA MODELU I OPIS ZMIENNYCH

Tworzony model ma na celu opisywać zależność między wybranymi zmiennymi i poparciem wyborczym dla PSL, PO i PIS w wyborach do Sejmu w 2007 i 2011 roku. Podstawową jednostką badawczą jest obszar powiatu (379 jednostek), przy czym dla niektórych zmiennych przyjęto, z powodu braku danych, wartości zmiennych dla subregionów (56 jednostek), w którym dany powiat się znajduje.

Wśród zmiennych endogenicznych w modelu znalazły się:

- 1) Poparcie wyborcze dla Prawa i Sprawiedliwości na poziomie powiatów w latach 2007 i 2011 (PIS).
- 2) Poparcie wyborcze dla Platformy Obywatelskiej na poziomie powiatów w latach 2007 i 2011 (PO).
- 3) Poparcie wyborcze dla Polskiego Stronnictwa Ludowego na poziomie powiatów w latach 2007 i 2011 (PSL).

Wśród zmiennych⁸ niezależnych wziętych pod uwagę w procesie budowy modeli regresyjnych znalazły się (kolejność przypadkowa)⁹:

- 1) Ogólny wskaźnik poziomu przestępczości dla podregionów, w których znajduje się dany powiat – stosunek liczby przestępstw w podregionie do liczby ludności w podregionie (oznaczenie CR).
Przyjęto wartości zmiennych dla roku poprzedzającego wybory, czyli dla roku 2006 i 2010.
- 2) Frekwencja wyborcza w powiatach w latach 2007 i 2011 (oznaczenie VT).
- 3) Urodzenia żywe na 1000 osób w 2007 i 2011 roku (FERTILITY). Przyjęto wartości zmiennych dla roku poprzedzającego wybory, czyli dla roku 2006 i 2010.

⁸ Źródła: Bank Danych Lokalnych GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, odczyt z dn. 20.02.2013, oraz PKW, <http://wybory2011.pkw.gov.pl/att/pl/000000.html#tabs-1>, odczyt z dn. 23.04.2013.

⁹ O ile nie wskazano wyraźnie, że jest inaczej, wszystkie zmienne wyrażone są w procentach, przy czym przyjęto przedział zmienności [0,1].

- 4) Stosunek liczby rozwodów do sumy liczby rozwodów i zawartych małżeństw w badanym roku w powiecie (DIV). Przyjęto dane dla powiatów, dla roku poprzedzającego wybory, czyli dla roku 2006 i 2010.
- 5) Stopa bezrobocia w powiatach (UNEMPL). Przyjęto wartości zmiennych dla roku poprzedzającego wybory, czyli dla roku 2006 i 2010.
- 6) Przeciętne miesięczne wynagrodzenia brutto w powiecie w relacji do średniej krajowej (Polska=100) (WAGES). Przyjęto wartości zmiennych dla roku poprzedzającego wybory, czyli dla roku 2006 i 2010.

Dodatkowo do modelu włączono dwie pomocnicze zmienne binarne oraz wskaźnik feminizacji:

- REGION – 1 dla powiatów wschodnich (Mazowsze, Ziemia Łódzka, Lubelszczyzna, Kielecczyzna i Podlasie - czyli dawny zabór rosyjski. Małopolska i Podkarpacie – czyli dawna Galicja), 0 dla powiatów znajdujących się na terenie pozostałych województw,
- It_2011–1 dla roku 2011, 0 dla roku 2007,
- FEM - wskaźnik feminizacji (liczba kobiet w relacji do ogólnej liczby mieszkańców powiatu).

W tabeli 2 przedstawiono symetryczną macierz korelacji zmiennych uwzględnionych w modelu (z wyjątkiem zmiennych zero-jedynkowych). Największy poziom korelacji liniowej występuje między zmiennymi FEM i DIV oraz między FEM i VT. Otrzymany wynik nie uzasadnia eliminacji niektórych zmiennych, gdyż korelacja między zmiennymi jest co najwyżej umiarkowana i wartość wskaźnika korelacji liniowej nie przekracza 0,75.

Tabela 2. Macierz korelacji zmiennych

	cr	vt	fertil~y	fem	unempl	wages	div_
cr	1.0000						
vt	0.2949	1.0000					
fertility	-0.1592	-0.1247	1.0000				
fem	0.3305	0.6039	-0.4057	1.0000			
unempl	0.0547	-0.4655	0.0007	-0.3199	1.0000		
wages	0.3003	0.4571	-0.1453	0.3924	-0.3849	1.0000	
div	0.5519	0.3778	-0.3344	0.6052	0.0553	0.3286	1.0000

Źródło: opracowanie własne.

POSTAĆ ANALITYCZNA MODELU

W celu rozwiązania przedstawionego w artykule problemu badawczego i sprawdzenia hipotez badawczych zastosowany został regresyjny model¹⁰ z efektami ustalonymi (*fixed effect*), którego parametry dla poszczególnych zmiennych objaśniających oszacowano, wykorzystując tzw. zero-jedynkową metodę najmniejszych kwadratów¹¹ (LSDV). Zero-jedynkowa metoda najmniejszych kwadratów jest metodą estymacji parametrów modeli regresyjnych budowanych na podstawie danych panelowych. Modelem regresyjnym dla danych panelowych, którego parametry szacujemy metodą LSDV, jest model regresji, który uzupełniony został o dodatkowe zmienne binarne. Dla przykładu, gdy analiza dotyczy dwóch momentów w czasie, wprowadzenie zmiennej zero-jedynkowej pozwala te momenty uwzględnić bezpośrednio w modelu. Przyjmując, że analiza dotyczy np. roku 2007 i 2011, możemy wartość 0 zmiennej binarnej przypisać do roku 2007, a wartość 1 do roku 2011. W ten sposób zmodyfikowany model pozwala uzyskać lepsze oszacowania poszukiwanych wartości parametrów. Postać analityczna modelu teoretycznego, uwzględniająca wszystkie wymienione wcześniej zmienne, jest następująca¹²:

¹⁰ Modele regresyjne to grupa modeli statystycznych danych w postaci równań matematycznych, które opisują stochastyczne związki między zmiennymi objaśnianymi (zależnymi) i objaśniającymi (niezależnymi). Modele regresyjne pozwalają lepiej zrozumieć zależności między zmiennymi, wyjaśnić kierunek zależności, ustalić siłę i charakter relacji między badanymi cechami itp. Popularne metody estymacji wartości parametrów modeli regresyjnych to metoda najmniejszych kwadratów MNK (*least-squares method*) oraz metoda największej wiarygodności (*maximum likelihood method*).

¹¹ Przystępne wprowadzenie do problematyki modeli regresyjnych i ich interpretacji można znaleźć np. w: M. Piłatowska, *Analiza współzależności zjawisk*, [w:] *Repetitorium ze statystyki*, Warszawa 2009, s. 92–129; D.M. Diez [et al.], *Introduction to Linear Regression*, [w:] *Openintro Statistics*, <http://www.openintro.org/stat/down/OpenIntroStatSecond.pdf>, odczyt z dn. 6.10.2013, s. 315–354; W. Mendenhall [et al.], *Introduction to Probability and Statistics*, 2013, s. 482–574. Na temat modeli regresyjnych stosowanych w analizie danych panelowych można przeczytać np. w: *Mikroekonometria*, red. M. Gruszczynski, Warszawa 2012, s. 267–308; C. Dougherty, *Introduction to Panel Data Models*, [w:] *Introduction to Econometrics*, 2011, s. 514–527; M. Pierzgałski, *Modele regresyjne w badaniu zachowań wyborczych*, „Athenaeum. Polskie Studia Politologiczne” 2012, vol. 36, s. 114–142.

¹² Podczas budowy poszczególnych modeli poparcia wyborczego i uwarunkowań frekwencji wyborczej wprowadzono niewielkie modyfikacje względem przedstawionej tu ogólnej postaci modelu.

$$y_{it} = \beta_0 + \beta_1 CR_{it} + \beta_2 VT_{it} + \beta_3 FERTILITY_{it} + \beta_4 DIV_{it} + \beta_6 FEM_{it} + \beta_7 WAGES_{it} + \beta_8 UNEMPL_{it} + \gamma_1 _It_2011_t + \delta_1 REGION_i + \mu_{it} \quad (1)$$

gdzie: $y_{it} = POit$ lub $PISit$ lub $PSLit$; $_It_2011^t = 1$, gdy $t = 2007$ i $_It_2011^t = 0$, gdy $t \neq 2007$; $REGION_i = 1$ dla powiatów wschodnich i $REGION_i = 0$ dla powiatów zachodniej Polski; μ_{it} – błąd czysto losowy.

W celu przeprowadzenia analizy zależności między wskazanymi zmiennymi zbudowano bazę danych zawierającą wszystkie obserwacje na tych zmiennych z lat 2007 i 2011 na poziomie powiatów¹³. Łączna liczba obserwacji na zmiennych dla poszczególnych partii politycznych jest więc równa $N \times T = 379 \times 2 = 758$. Baza danych jest zbyt rozbudowana, aby można było ją bezpośrednio włączyć do artykułu¹⁴. Ponieważ w badaniu uwzględniono wszystkie możliwe obserwacje, analiza ma charakter badania pełnego. W tym przypadku elementy badanej próby są tożsame z całą populacją jednostek badawczych (powiaty).

Przed rozpoczęciem estymacji parametrów modelu wygenerowano zmienną zero-jedynkową. Pozwoliła ona uwzględnić efekty ustalone w wymiarze czasowym (*time fixed effects* – np. zmiana stanu gospodarki spowodowana kryzysem gospodarczym, zmiana formatu systemu partyjnego), ponieważ istnieje możliwość, że dla poszczególnych okresów mamy dwie różne linie regresji – różnią się wyrazem wolnym. Dodatkowo wprowadzona została zmienna binarna *REGION*, która różnicuje badane jednostki (powiaty) w wymiarze Polska B – Polska A¹⁵.

Model objaśnia poparcie wyborcze dla trzech partii politycznych, a mianowicie dla PSL, PO i PIS, w wyborach 2007 i 2011. Ogólnie, model opisuje liniowe związki między zmiennymi, ale w jednym, szczególnie uzasadnionym przypadku wprowadzono do modelu modyfikację, która pozwoliła na opis zależności nieliniowej. Ponadto w artykule przedstawiono jedynie ostateczne postaci modeli, po eliminacji zmiennych, których wpływ na zmienne objaśniane okazał się dyskusyjny.

¹³ Źródłem danych były strona internetowa Państwowej Komisji Wyborczej, <http://pkw.gov.pl/> oraz Bank Danych Lokalnych GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks.

¹⁴ Zbiorem danych w pliku arkusza kalkulacyjnego będących podstawą do budowy modeli dysponuje autor.

¹⁵ Do powiatów Polski B zaliczono te, które znalazły się w granicach regionów, w których w wyborach prezydenckich 2010 zwyciężył J. Kaczyński, a mianowicie – Mazowsze, Ziemia Łódzka, Lubelszczyzna, Kielecczyzna i Podlasie – czyli dawny zabór rosyjski. Małopolska i Podkarpacie – czyli dawna Galicja.

ESTYMACJA ORAZ WERYFIKACJA
STATYSTYCZNA MODELI

W tej części artykułu przedstawiono specyfikację modeli regresyjnych zbudowanych w celu weryfikacji przedstawionych hipotez badawczych (estymację wartości parametrów modeli przeprowadzono metodą LSDV). Do równań modeli włączono jedynie te zmienne, które po analizie diagramów rozproszenia, ilustrujących charakter zależności między zmiennymi objaśnianymi i objaśniającymi, oraz po ocenie ich poziomu istotności można było uznać za czynniki faktycznie oddziałujące na poziom zmiennych zależnych.

Poniżej zaprezentowane zostaną kolejno modele zależności między poparciem wyborczym a zmiennymi społeczno-gospodarczymi związanymi z poziomem kapitału społecznego dla PO, PiS i PSL. Na koniec przedstawiony zostanie dodatkowo model opisujący wpływ badanych zmiennych na frekwencję wyborczą.

Wyniki estymacji modelu dla PO przedstawiono w tabeli 3. W modelu dla PO najbardziej istotna okazuje się zmienna DIV. Ponadto wysokie wartości statystyki

Tabela 3. Model poparcia wyborczego dla PO

Source	SS	df	MS	Number of obs = 758		
-----+-----				F(5, 752) = 219.39		
Model	5.77059223	5	1.15411845	Prob > F = 0.0000		
Residual	3.95589305	752	.005260496	R-squared = 0.5933		
-----+-----				Adj R-squared = 0.5906		
Total	9.72648528	757	.012848726	Root MSE = .07253		
-----+-----						
po	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
cr	4.015004	.4102728	9.79	0.000	3.209587	4.82042
vt	.1595145	.0499467	3.19	0.001	.0614629	.2575662
fem	1.335445	.4427526	3.02	0.003	.4662672	2.204623
div	.7281169	.052992	13.74	0.000	.624087	.8321467
egion	-.0290765	.0060245	-4.83	0.000	-.0409034	-.0172496
_cons	-.6609626	.2110773	-3.13	0.002	-1.075333	-.2465918
-----+-----						

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

t^{16} uzyskano dla zmiennych CR i VT. Do modelu włączono też zmienną FEM, która uzyskała akceptowalny, ale stosunkowo niski poziom istotności, co jest jednak spowodowane relatywnie wysokim poziomem skorelowania tej zmiennej ze zmienną DIV.

Z modelu usunięto zmienne UNEMPL, WAGES, FERTILITY i zero-jedynkowa zmienną $_It_2011$, gdyż analiza nie wykazała istotnego związku między tymi zmiennymi a poparciem wyborczym PO.

Równanie modelu jest następujące (oceny parametrów podano z zaokrągleniem do trzech miejsc po przecinku):

$$PO_{it} = -0,661 + 4,015CR_{it} + 0,160VT_{it} + 0,728DIV_{it} + 1,335FEM_{it} - 0,029REGION_i \quad (2)$$

Uzyskano stosunkowo wysoki poziom wskaźnika determinacji, który wynosi w przybliżeniu 0,59, co oznacza, że model wyjaśnia ok. 59% wariacji poparcia dla PO. Dodatkowo wartość odchylenia standardowego resztowego modelu wynosi w przybliżeniu 0,07, a zatem szacując względny poziom poparcia wyborczego dla PO za pomocą omawianego modelu, mylimy się przeciętnie o 7 punktów procentowych (p.p.).

Model dla Prawa i Sprawiedliwości przedstawiony został w tabeli 4. W porównaniu z modelem dla PO stopień objaśnienia wariacji zmiennej objaśnianej przez model jest niższy i wynosi w przybliżeniu 0,44. Błąd standardowy modelu przyjmuje podobną wartość jak dla PO i wynosi blisko 0,74. W przypadku modelu dla PIS nieistotne statystycznie okazują się zmienne UNEMPL, WAGES i FERTILITY.

Otrzymano następujący model:

$$PIS_{it} = 0,456 - 2,661CR_{it} - 1,406VT_{2it} + 1,858VT_{it} - 0,677DIV_{it} + \\ -0,937FEM_{it} - 0,024_It_2011 + 0,036REGION_i \quad (3)$$

¹⁶ Statystyka testowa t jest stosowana podczas weryfikacji statystycznej modeli regresyjnych. Jej wartość pomaga ocenić, czy dana zmienna, dla której wyznaczono wartość tej statystyki, jest statystycznie istotna, czyli czy możemy wnioskować, że dana zmienna wpływa na zmienną objaśnianą. Tzw. wartości krytyczne, z którymi porównujemy otrzymaną wartość statystyki t i następnie oceniamy istotność zmiennej, odczytujemy z tablic rozkładu t -Studenta [Stosując modele regresji liniowej przyjmujemy, że estymatory parametrów (współczynników) równań modelu regresji mają rozkład t -Studenta]. Ogólnie im wyższa wartość statystyki t , tym większe prawdopodobieństwo, że zmienne są istotne.

Tabela 4. Model poparcia wyborczego dla PIS

Source	SS	df	MS	Number of obs = 758		
-----+-----				F(7,750) = 85.27		
Model	3.31028315	7	.472897592	Prob > F = 0.0000		
Residual	4.15959172	750	.005546122	R-squared = 0.4432		
-----+-----				Adj R-squared = 0.4380		
Total	7.46987486	757	.009867734	Root MSE = .07447		
-----+-----						
pis	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
cr	-2.660507	.4283036	-6.21	0.000	-3.501323	-1.81969
vt2	-1.405528	.4560042	-3.08	0.002	-2.300725	-.510332
vt	1.858339	.4525934	4.11	0.000	.9698387	2.74684
fem	-.9366289	.4945704	-1.89	0.059	-1.907536	.0342781
div	-.6769652	.0550052	-12.31	0.000	-.7849477	-.5689828
region	.0358659	.0061859	5.80	0.000	.0237221	.0480097
t_2011	-.0242499	.006299	-3.85	0.000	-.0366157	-.0118841
_cons	.4556986	.2728525	1.67	0.095	-.0799468	.991344
-----+-----						

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

Do modelu dla PIS wprowadzono dodatkowo zmienne VT2 (kwadrat zmiennej VT) w celu ukazania parabolicznej zależności między zmiennymi VT i PIS. Model dla PSL pokazano w tabeli 5.

W modelu dla PSL, podobnie jak w przypadku modeli dla PO i PIS, istotne okazały się zmienne CR, VT, DIV, FEM oraz dodatkowo zmienna *_It_2011* i zmienna REGION.

Postać analityczna modelu jest następująca:

$$PSL_{it} = 0,895 - 1,606CR_{it} - 0,381VT_{it} - 0,335DIV_{it} - 0,897FEM_{it} + \\ -0,042_It_2011_t - 0,021REGION_i \quad (4)$$

Model dla PSL w najmniejszym stopniu spośród trzech zbudowanych modeli objaśniających poparcie wyborcze wyjaśnia zmienność poparcia wyborczego. R2 jest równy w przybliżeniu 0,35¹⁷. Błąd standardowy modelu jest w tym przypadku równy ok. 0,07.

Tabela 5. Model poparcia wyborczego dla PSL

Source	SS	df	MS	Number of obs = 758		
				F(6, 751) = 67.88		
Model	2.13620769	6	.356034615	Prob > F = 0.0000		
Residual	3.9388592	751	.005244819	R-squared = 0.3516		
				Adj R-squared = 0.3465		
Total	6.07506689	757	.008025187	Root MSE = .07242		
psl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cr	-1.605753	.4130703	-3.89	0.000	-2.416663	-.7948434
vt	-.3808507	.0556484	-6.84	0.000	-.4900955	-.2716059
fem	-.8967422	.4753114	-1.89	0.060	-1.829839	.0363549
div_	-.3348331	.0534655	-6.26	0.000	-.4397927	-.2298734
It_2011	-.042203	.0060987	-6.92	0.000	-.0541755	-.0302306
region	.0208469	.0060156	3.47	0.001	.0090376	.0326562
_cons	.895129	.2225602	4.02	0.000	.4582149	1.332043

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

¹⁷ Współczynnik R2 jest miarą dobroci dopasowania modelu regresji liniowej do danych empirycznych i przyjmuje wartości z przedziału [0,1].

Tabela 6. Model opisujący uwarunkowania frekwencji wyborczej (VT)

Source	SS	df	MS	Number of obs = 758		
-----+-----				F(7,750) = 278.01		
Model	2.44312553	7	.349017933	Prob > F = 0.0000		
Residual	.94155925	750	.001255412	R-squared = 0.7218		
-----+-----				Adj R-squared = 0.7192		
Total	3.38468478	757	.004471182	Root MSE = .03543		

vt	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]
-----+-----						
fertility	.0162085	.0011662	13.90	0.000	.0139191	.0184979
unempl	-.326146	.0235798	-13.83	0.000	-.3724362	-.2798557
wages	.0930996	.0122142	7.62	0.000	.0691215	.1170776
fem	4.161612	.2191731	18.99	0.000	3.731346	4.591877
div	.0355354	.0253204	1.40	0.161	-.0141718	.0852426
_It_2011	-.0516762	.0026835	-19.26	0.000	-.0569443	-.046408
region	.0052342	.0029449	1.78	0.076	-.000547	.0110154
_cons	-1.82786	.1148347	-15.92	0.000	-2.053296	-1.602424

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

Dodatkowo w tabeli 6 przedstawiono model dla frekwencji wyborczej. Równanie modelu jest następujące:

$$VT_{it} = -1,828 + 0,016FERTILITY_{it} + 4,162FEM_{it} + 0,036DIV_{it} + 0,093WAGES_{it} - 0,326UNEMPL_{it} - 0,052_It_2011_t + 0,005REGION_i \quad (5)$$

W tym przypadku istotnie statystycznie okazały się zmienne FERTILITY, FEM, DIV, WAGES, UNEMPL, _It_2011 i REGION. Należy podkreślić, że uzyskano wysoki poziom objaśnienia, który wynosi ok. 0,72. Zadowolająca jest też wartość błędu modelu i równa w przybliżeniu 0,035.

Podsumowując, efektem przeprowadzonych badań są modele, których jakość, uwzględniając specyfikę pola badawczego, można uważać za zadowalającą. W szczególności model dla frekwencji wyborczej cechuje wysoki poziom obja-

śnienia wariancji. Stosunkowo wysoką wartość R^2 otrzymano dla zmiennych PO, PIS. Co ważne, większość analizowanych zmiennych okazała się istotna statystycznie.

W następnej części artykułu przedstawiona zostanie weryfikacja merytoryczna modeli.

WERYFIKACJA MERYTORYCZNA MODELI – WNIOSKI

Platforma Obywatelska

W przypadku modelu dla PO interesujące są wyniki dotyczące zmiennej CR. Otóż ujawnia się pozytywny związek między poziomem przestępczości a poparciem wyborczym PO. Zmienną CR cechuje wysoki poziom istotności, na co wskazuje wysoka wartość statystyki t -studenta. Z modelu wynika, że przy wzroście wskaźnika CR o jednostkę (jeden punkt procentowy) należy oczekiwać wzrostu poparcia wyborczego PO średnio o ok. 4 p.p. (przy założeniu *ceteris paribus* odnośnie do pozostałych zmiennych uwzględnionych w danym modelu). Jak można wyjaśnić tę zależność?

Wynik badania wskazuje, że w podregionach, w których poziom przestępczości jest wyższy, poparcie dla PO jest większe. Być może, jest to po prostu związane z faktem, że w podregionach lepiej rozwiniętych gospodarczo, zarówno poparcie dla względnie liberalnej gospodarczo PO, jak również poziom przestępczości są zwykle większe. Niemniej warto przyjrzeć się dokładniej omawianej zależności. Przedstawiono ją na rysunku 1, przy czym wprowadzono dodatkową zmienną binarną, która umożliwiła wykreślenie diagramu z pominięciem największych i najlepiej rozwiniętych gospodarczo miast, które samodzielnie tworzą strukturę subregionów, a w których poziom przestępczości jest zwykle relatywnie wysoki.

Można tę prawidłowość poprzeć dowodem w postaci prostego modelu regresji logistycznej. Wprowadźmy nową zmienną METRO, która przyjmuje wartość 1 wtedy, gdy subregionem jest samodzielnie duże miasto w Polsce, dla pozostałych przypadków przyjmuje wartość 0.

Z modelu przedstawionego w tabeli 6 wynika, że prawdopodobieństwo wystąpienia wyższego poziomu przestępczości jest większe dla dużych miast.

Tabela 7. Model logitowy zależności między zmiennymi METRO i CR

Logistic regression		Number of obs	=	758		
		LR chi2(1)	=	43.61		
		Prob > chi2	=	0.0000		
Log likelihood = -39.847119		Pseudo R2	=	0.3537		

metro	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
cr	192.313	32.26713	5.96	0.000	129.0706	255.5554
_cons	-11.25855	1.490847	-7.55	0.000	-14.18056	-8.336546

Źródło: opracowanie własne na podstawie danych GUS: http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, odczyt z dn. 5.10.2012.

Analizując rysunek 1, można zauważyć, że zależność między zmiennymi wynik PO i CR jest dodatnia, przy czym wraz ze wzrostem poziomu przestępczości obserwujemy coraz wolniejszy wzrost poparcia dla PO. Mamy tutaj zatem *de facto* przykład zależności o charakterze nieliniowym. W przypadku tej zmiennej można by zaproponować włączenie do modelu jej logarytmu.

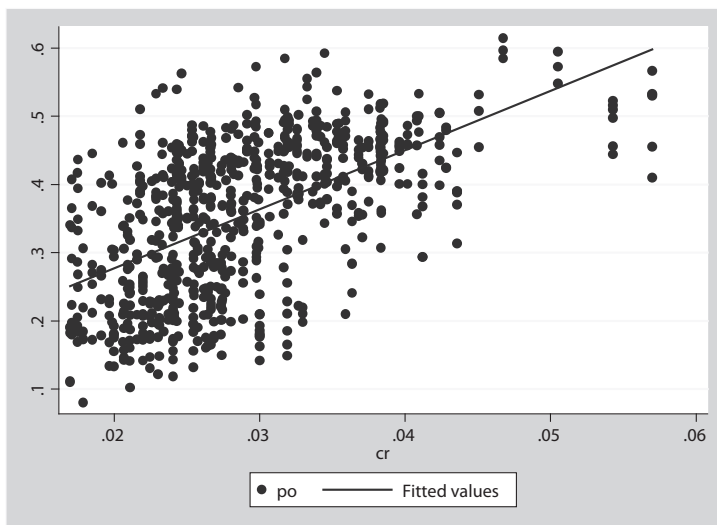
Dla porównania, gdyby zbudowano model opisujący zależność między PO i CR, uwzględniający jedynie duże miasta tworzące samodzielnie podregiony, to omawianą zależność jednokierunkową również można zaobserwować. Model przedstawiony na rysunku 2 potwierdza istnienie dodatniej zależności między zmiennymi PO i CR.

Odnosząc się do zmiennej VT, trzeba zauważyć, że pozytywnie weryfikuje się hipoteza głosząca, iż wzrost frekwencji wyborczej w powiatach jest dodatnio skorelowany ze wzrostem poparcia wyborczego dla PO. Innymi słowy, Platformie Obywatelskiej powinno zależeć na wyższej frekwencji. Jest to zgodne z intuicją, która wskazuje, że potencjalny elektorat PO jest, w porównaniu ze zdyscyplinowanymi wyborcami PIS, bardziej pasywny.

Okazuje się, że dla zmiennych: poparcie dla PO oraz wzrost natężenia rozwodów obserwujemy zależność jednokierunkową.

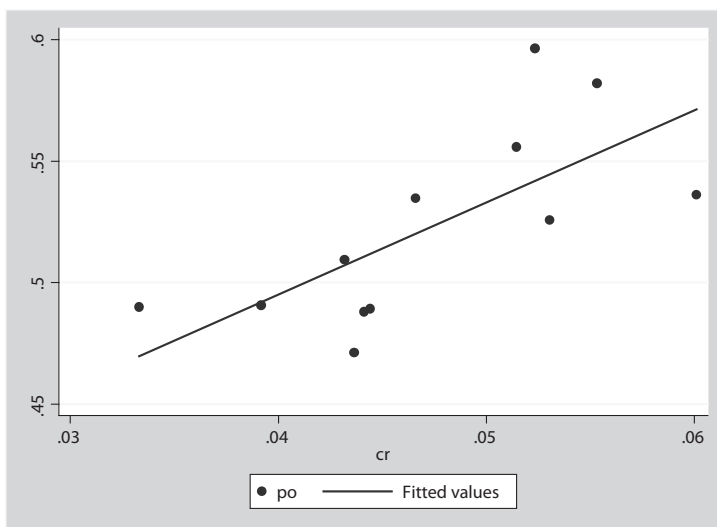
Ponadto na PO chętniej głosuje się w powiatach o wyższym stopniu feminizacji.

Model uprawdopodobnia również, że w powiatach tzw. Polski B poparcie dla PO jest mniejsze – ujemna wartość parametru przy zmiennej REGION.



Rysunek 1: Model dla zależności między poparciem wyborczym dla PO a poziomem przestępczości z pominięciem dużych miast tworzących samodzielnie strukturę subregionu (Warszawa, Kraków, Łódź, Wrocław, Poznań, Szczecin)

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.



Rysunek 2: Model dla zależności między poparciem wyborczym dla PO a poziomem przestępczości w dużych miastach tworzących samodzielnie strukturę subregionu

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

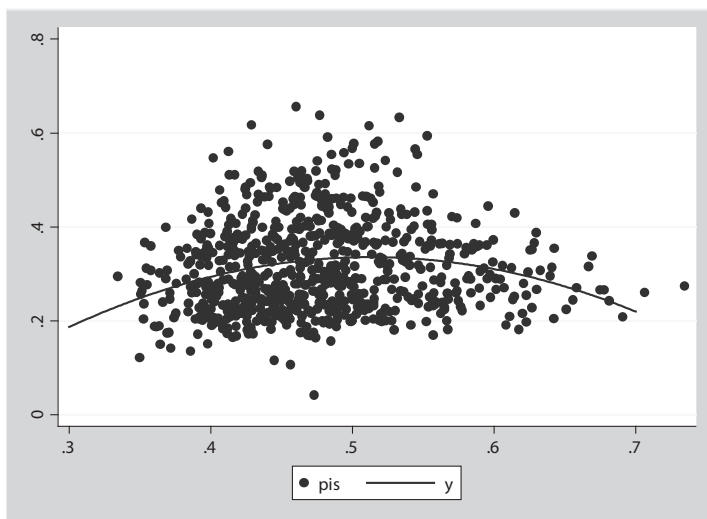
Prawo i Sprawiedliwość

Odnosząc się do zmiennej CR w przypadku PiS, ujawnia się między zmiennymi zależność odwrotna, ale nie jest ona już tak wyraźna jak w przypadku PO.

W przypadku Prawa i Sprawiedliwości ciekawy jest wynik dotyczący zmiennej VT. Otóż okazuje się, że na podstawie ocen parametrów modelu dla PIS można wnioskować, że wzrost poziomu frekwencji wyborczej sprzyja poparciu dla Prawa i Sprawiedliwości, ale tylko do pewnego momentu, ponieważ relacja między zmiennymi przypomina zależność paraboliczną – rysunek 3.

Na podstawie modelu wykreślonego na rysunku 3 można łatwo wyznaczyć maksimum równania regresji dla relacji PIS i VT. Otóż okazuje się, że znajduje się ono w przybliżeniu w punkcie (0,51, 0,34). Można zatem stwierdzić, że optymalny poziom frekwencji wyborczej w powiatach z punktu widzenia PIS wynosi ok. 51%, przy którym poparcie wyborcze dla PiS w roku 2007 i 2011 wyniosłoby ok. 34%.

Intuicja wskazuje, że natężenie rozwodów będzie ujemnie skorelowane z poziomem poparcia wyborczego dla PIS. Równanie modelu potwierdza to przypuszczenie, gdyż współczynnik przy zmiennej DIV jest mniejszy od zera i statystycznie istotny.



Rysunek 3: Model przedstawiający zależność między zmiennymi PIS i VT

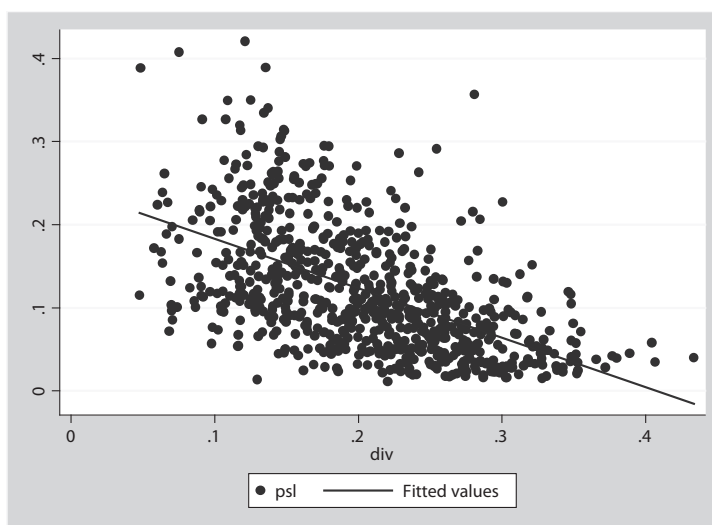
Źródło: opracowanie własne na podstawie danych: PKW, <http://pkw.gov.pl/> odczyt z dn. 5.07.2013.

Model dla PIS wskazuje, że im większy jest udział kobiet w populacji powiatów, tym mniejsze jest poparcie udzielane PiS – ujemny parametr przy zmiennej FEM.

Model potwierdza również empirycznie obserwowalny fakt, że w powiatach znajdujących się na obszarze tzw. Polski B relatywnie dużym poparciem cieszy się właśnie Prawo i Sprawiedliwość. Dodatni współczynnik przy zmiennej REGION, który jest równy w przybliżeniu 0,036, informuje, że w powiatach wschodniej Polski należy oczekiwać, że poparcie udzielone PIS będzie przeciętnie większe o ok. 3,6 p.p. w porównaniu z powiatami zachodnimi (*ceteris paribus*).

Polskie Stronnictwo Ludowe

Dla PSL zmienna CR wykazuje niezbyt wysoki poziom istotności, a ocena odpowiedniego parametru modelu wykazuje negatywny związek między zmiennymi PSL i CR, co dowodzi, że większy poziom przestępczości jest negatywnie skorelowany z poziomem poparcia PSL.



Rysunek 4: Model przedstawiający zależność między zmiennymi PSL i DIV¹⁸

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

¹⁸ Ze zbioru danych wyłączono kilka obserwacji nietypowych (*outliers*) w celu zwiększenia czytelności diagramu.

W przypadku zmiennej VT należy odnotować, że dla PSL wzrost VT jest dla poziomu poparcia tej partii politycznej niekorzystny, co można uzasadnić, odwołując się do specyfiki elektoratu tej partii. Badanie najprawdopodobniej potwierdza przypuszczenie, że elektorat PSL cechuje stosunkowo stały poziom frekwencji wyborczej, dlatego ogólny wzrost poziomu VT przyczynia się do relatywnego pogorszenia wyniku wyborczego PSL, gdyż głosujący, którzy ostatecznie przyczynią się do osiągnięcia przez powiat relatywnie wysokiej frekwencji (np. grupa wyborców niezdecydowanych, czy wziąć udział w głosowaniu), zwykle nie oddają swoich głosów na PSL.

W modelu opisującym poparcie dla PSL jednoznacznie istotna okazuje się zmienna DIV, której oszacowany parametr jest ujemny, i tym samym wskazuje, że względny wzrost liczby rozwodów związany jest ze spadkiem poparcia wyborczego dla PSL.

Odnosząc się do zmiennej FEM, odpowiednia ocena wartości parametru dowodzi, że istnieje zależność odwrotna między tą zmienną a wielkością poparcia wyborczego dla PSL, a zatem można sądzić, że w powiatach o mniejszym poziomie feminizacji chętniej głosuje się na Polskie Stronnictwo Ludowe.

Frekwencja wyborcza

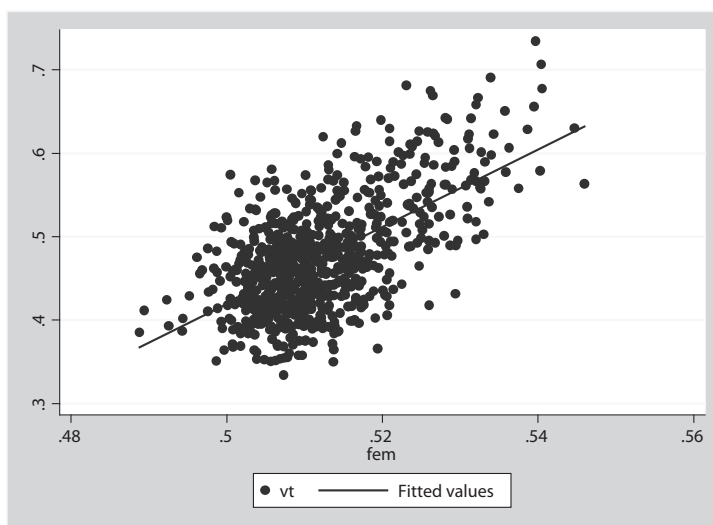
Na wstępie warto zauważyć, że wartość estymatora parametru istotnej zmiennej *_It_2011* potwierdza spadek frekwencji wyborczej w elekcji 2011 roku.

Jakie inne ciekawe wnioski wynikają z analizy modelu dla VT? Otóż model dowodzi, że **frekwencja wyborcza w powiatach o wyższym wskaźniku feminizacji jest wyższa**. Czy z tego wynika, że kobiety częściej uczestniczą w wyborach w porównaniu z mężczyznami? Niektóre badania wskazują, że w Polsce w wyborach częściej uczestniczą mężczyźni¹⁹.

Niemniej istnieją też źródła, które dowodzą, że to wśród kobiet frekwencja wyborcza jest wyższa²⁰.

¹⁹ M. Cześnik, *Partycypacja wyborcza Polaków*, 2009, <http://www.isp.org.pl/files/20145849250174351001263374709.pdf>, odczyt z dn. 8.12.2012.

²⁰ Zob. np. *Frekwencja wyborcza między 53,15 a 55,3 proc.*, http://fakty.interia.pl/raport/wybor-y_2007/news/frekwencja-wyborcza-miedzy-5315-a-553-proc,997349,4727, odczyt z dn. 28.11.2012; *Sondaż PBS DGA: PO – 44 proc., PiS – 30 proc., LiD – 12 proc.*, „Gazeta.pl”, <http://wiadomosci.gazeta.pl/wiadomosci/1,114873,4599291.html>, odczyt z dn. 28.11.2012.



Rysunek 5: Model przedstawiający zależność między zmiennymi VT i FEM

Źródło: opracowanie własne na podstawie danych: GUS, http://www.stat.gov.pl/bdl/app/strona.html?p_name=indeks, PKW, Online: <http://pkw.gov.pl/>, odczyt z dn. 5.07.2013.

Z czego biorą się te rozbieżności? Zdaje się, że mogą one być związane z metodologią prowadzonych badań partycypacji obywateli – wszystkie dane, które są dostępne na temat rozkładu głosów wśród uprawnionych do uczestnictwa w wyborach, uzyskiwane są metodą sondażową²¹.

Wyniki sondażu *exit poll* przeprowadzonego w dniu wyborów przez TNS OBOP dla TVP wskazały, że głosowało 55,6% mężczyzn oraz 48,4% kobiet. Wyniki tych badań dotyczące frekwencji wyborczej dla roku 2007 mogą jednak być błędne. Dla przykładu, według sondażu PBS DGA wykonanego na zlecenie Telewizji TVN w dniu wyborów parlamentarnych 2007 frekwencja wśród kobiet wyniosła 58,3%, a w populacji mężczyzn 53%²² – było to badanie typu *exit poll*. Rozbieżność wyników badań sondażowych jest więc znaczna, co skłania do dalszych badań²³. Odnośnie do roku 2011, badanie *exit poll* przeprowadzone

²¹ *Panie poszły do urn*, „Nowiny.pl”, <http://www.nowiny.pl/78172-panie-poszly-do-urn.html>, odczyt z dn. 29.11.2012.

²² *Sondaż PBS DGA: PO – 44 proc., PiS – 30 proc., LiD – 12 proc.*, „Gazeta.pl”, op.cit.

²³ Państwowa Komisja Wyborcza nie gromadzi danych dotyczących frekwencji wyborczej dla płci, gdyż w świetle obowiązującego w Polsce prawa w przypadku wyborów do Sejmu archiwizowane

przez TNS OBOP dla TVP wykazały, że głosowało 50,1% mężczyzn oraz 47,6% kobiet²⁴. PBS sondażu *exit poll* w 2011 roku nie przeprowadzał.

Rysunek 5 pokazuje, że dodatnia korelacja między VT i FEM jest wyraźna, co potwierdza wyniki, które uzyskał PBS DGA w 2007 roku. Uzyskany wynik przekonuje, że w powiatach, w których odsetek kobiet jest wyższy, wyższa jest również frekwencja. Oparte na danych zagregowanych wyniki analiz mogą ponadto uprawdopodobniać wyniki badań, które wskazują, że kobiety częściej biorą udział w głosowaniach, ale nie można w tym przypadku zapomnieć o możliwości popełnienia błędu wynikającego z wnioskowania opartego na tzw. danych ekologicznych²⁵ (*ecological inference fallacy*).

Jeśli chodzi o stopę bezrobocia, to potwierdza się przypuszczenie, że jej wzrost negatywnie oddziałuje na poziom frekwencji.

Model dowodzi także, że przeciętnie poziom frekwencji wyborczej nie różnicuje powiatów znajdujących się na terytorium tzw. Polski B i powiatów Polski A.

Poza tym na poziom frekwencji wyborczej ma korzystny wpływ wyższy przyrost naturalny.

Wyniki badań ujawnione w artykule dostarczają wiedzę, która pozwala na weryfikację, z pewnym prawdopodobieństwem popełnienia błędu, przedstawionych w artykule hipotez badawczych.

są tylko dane dotyczące liczby głosów ważnych i głosów ważnych oddanych na każdą listę kandydatów oraz liczba głosów ważnych oddanych na poszczególnych kandydatów z każdej z list. Zob. *Ustawa z dnia 5 stycznia 2011 – Kodeks wyborczy*, „Dziennik Ustaw” 2011, nr 21, poz. 112 z późn. zm.

²⁴ Porównując ten wynik z wynikiem sondażu, który TNS OBOP przeprowadził w 2007 roku, można zauważyć znaczne zmniejszenie różnicy między frekwencją wśród kobiet i wśród mężczyzn.

²⁵ Błąd ekologiczny może pojawić się, gdy na podstawie danych zagregowanych (analiza danych może być prowadzona na danych indywidualnych lub zagregowanych) wyciągamy wnioski na temat jednostek, które w skład danych agregatów wchodzi. Np. gdy z faktu, że korelacja między poziomem frekwencji wyborczej i odsetkiem kobiet w powiatach (dane zagregowane na poziomie powiatów) jest pozytywna, będziemy wnioskować, że kobiety chętniej uczestniczą w wyborach (poziom jednostek – dane indywidualne). Takie wnioskowanie może być nieprawidłowe. W tym przypadku można jedynie niezawodnie wnioskować, że w powiatach o większej liczbie kobiet poziom frekwencji wyborczej jest przeciętnie większy. Innymi słowy, operując na danych zagregowanych na poziomie np. gmin, wnioski, co do zasady, powinny dotyczyć gmin, a nie np. elementów składowych danego agregatu. O problemie wnioskowania na podstawie danych zagregowanych i na temat tzw. błędów ekologicznych zob. np. w: G. King, *A Solution to the Ecological Inference Problem*, Princeton 1997, s. 1–27.

PODSUMOWANIE

Celem artykułu było wyjaśnienie charakteru zależności między zachowaniami wyborczymi Polaków i wybranymi zmiennymi społeczno-gospodarczymi. Podstawową jednostką badawczą był obszar powiatu, a modele zostały zbudowane na podstawie 758 obserwacji na każdej spośród analizowanych zmiennych. Badanie dotyczyło wyborów parlamentarnych w 2007 i 2011 roku.

Hipoteza 1 nie potwierdziła się. Otóż okazuje się, że w przypadku zmiennej DIV, której wzrost wpływa negatywnie na poziom kapitału społecznego, w modelu opisującym poparcie wyborcze dla PO otrzymano dodatnią wartość estymatora parametru przy badanej zmiennej. Podobnie jest w przypadku poziomu przestępczości, którego wzrost przyczynia się do zwiększenia zasobów negatywnego kapitału społecznego. Model dla PO wskazuje korelację pozytywną z poziomem poparcia wyborczego.

Hipoteza 2 została zweryfikowana pozytywnie, jednakże z jednym zastrzeżeniem. Otóż w przypadku Prawa i Sprawiedliwości zależność między frekwencją wyborczą i poziomem poparcia wyborczego została opisana wielomianem drugiego stopnia, który lepiej w porównaniu z funkcją liniową opisał omawianą zależność. Z tego powodu nie można jednoznacznie określić kierunku zależności między zmiennymi. Jednak mniej więcej do pewnego poziomu frekwencji wyborczej, wynoszącego ok. 51%, można stwierdzić, że im większa VT, tym wyższy poziom poparcia PIS, natomiast powyżej 51% poparcie dla PiS stopniowo maleje.

Przedstawione badania falsyfikują hipotezę 3. Okazuje się, że wzrost względnej liczby kobiet w powiecie prowadzi do zwiększenia frekwencji wyborczej. Powstaje jeszcze pytanie, czy ten wniosek, który został osiągnięty wskutek analizy agregatów, można przenieść na zachowanie jednostek. Innymi słowy, czy można twierdzić na podstawie uzyskanego wyniku, że kobiety chętniej od mężczyzn uczestniczą w wyborach parlamentarnych w Polsce. Takie rozumowanie może, jak wspomniano, prowadzić do popełnienia w tym przypadku błędu wynikającego z wnioskania na tzw. danych ekologicznych. Wprawdzie liczne badania dowodzą, że kobiety stosunkowo rzadziej biorą udział w głosowaniach, ale wyniki zaprezentowane w artykule, przy założeniu, że nie wystąpił błąd ekologiczny, stanowią argument przemawiający za prawdziwością ustaleń tych sondaży (np. sondaż PBS dla telewizji TVN przeprowadzony w dniu wyborów parlamentarnych 2007), których wyniki są w zgodzie z wnioskami płynącymi z analizy modelu regresyjnego, który zbudowano na potrzeby artykułu.

Z modeli przedstawionych w artykule wynika, że rozważane w tekście czynniki ekonomiczne (UNEMPL i WAGES) nie są istotnie powiązane z poziomem poparcia wyborczego udzielonego badanym partiom politycznym (hipoteza 4). Analiza przedstawiona w artykule dostarcza argumenty na rzecz hipotezy, która zakłada, że rywalizacja wyborcza w Polsce warunkowana jest przede wszystkim czynnikami kulturowo-obyczajowymi i demograficznymi.

W zakończeniu warto podsumować analizowane zależności między zmiennymi, które ujawniono w modelach.

Tabela A. Podsumowanie ujawnionych w modelach zależności między zmiennymi

Zmienna objaśniana	Zmienne objaśniające								
	CR	VT	FERTILITY	DIV	FEM	UNEMPL	WAGES	REGION	_It_2011
PO	+	+	<i>n</i>	+	+	<i>n</i>	<i>n</i>	-	<i>n</i>
PIS	-	+/-	<i>n</i>	-	-	<i>n</i>	<i>n</i>	+	-
PSL	-	-	<i>n</i>	-	-	<i>n</i>	<i>n</i>	+	-
VT	<i>n</i>		+	+	+	-	+	+	-

Źródło: opracowanie własne: „+” oznacza związek jednokierunkowy, „-” oznacza relację odwrotną między zmiennymi, „*n*” – zmienna jest nieistotna.

Z tabeli A wynika m.in., że zbiór zmiennych, które istotnie oddziałują na poziom poparcia wyborczego nie zmienia się w zależności od badanej partii politycznej. To stwierdzenie nie dotyczy tylko zmiennych zero-jedynkowych.

Ponadto z tabeli A wynika znaczne podobieństwo PIS i PSL z punktu widzenia charakteru wpływu zmiennych niezależnych włączonych do modelu na wyniki wyborcze tych partii. W szczególności można stwierdzić, że poparcie wyborcze dla PSL i PIS okazuje się pozytywnie skorelowane. Potwierdza tę obserwację tabela B.

Tabela B. Symetryczna macierz korelacji między zmiennymi PSL, PO i PIS

	PSL	PO	PIS
PSL	1		
PO	-0,640	1	
PIS	0,245	-0,698	1

Źródło: opracowanie własne.

Dane przedstawione w tabeli B dowodzą, że wyborcy PSL, chętniej niż PO, popierają PIS, co jak się wydaje, wynika z podobnych preferencji politycznych charakteryzujących te grupy wyborców. Wydaje się, że koalicja rządowa, w której skład wchodzi PIS i PSL, jest bardziej naturalna.

Poza tym tabela A może stanowić pewną wskazówkę dla badanych partii politycznych. Dla przykładu, badania wykazały, że wzrost poziomu frekwencji wyborczej jest niekorzystny dla PIS i PSL, natomiast sprzyja Platformie Obywatelskiej – widoczne jest to w kolumnie VT tabeli A. Z tabeli A możemy też odczytać, że w powiatach, w których rośnie liczba rozwodów, zwiększa się też poparcie wyborcze PO, natomiast pogarsza się wynik wyborczy PSL i PIS.

Jeśli chodzi o poziom poparcia wyborczego, przedstawione modele nie ujawniły istotnego wpływu czynników ekonomicznych. W przypadku poparcia wyborczego silniej zaznacza się wpływ czynników natury społeczno-kulturowej oraz poziom frekwencji wyborczej. Natomiast frekwencja wyborcza jest związana również z czynnikami *stricte* ekonomicznymi, takimi jak stopa bezrobocia.

Przedstawione badania nie wyczerpują rozważanego problemu badawczego, ponieważ zarówno liczba zmiennych, które można by wziąć pod uwagę, nie jest pełna, jak również okres badania nie jest wystarczająco długi. Niemniej uzyskane wyniki pogłębiają wiedzę o naturze zachowań wyborczych w Polsce.